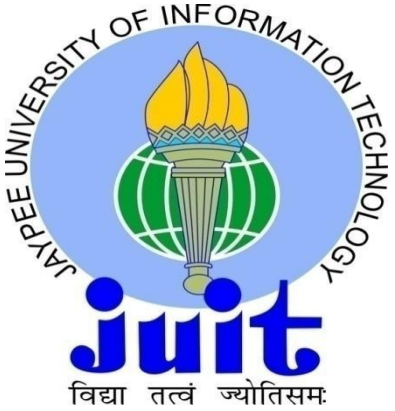
# Comprehensive Knowledge Generation for Clinical, Radiology and Genetics of Interstitial Lung Diseases (ILDs)

*Synopsis submitted in fulfillment of the requirment for the Degree of*

**DOCTOR OF PHILOSOPHY**

IN

**BIOINFORMATICS**




BY

**Smriti Mishra**
Enrollment No. 136505

Under the Supervision of

**Dr. Jayashree Ramana**

Department of Biotechnology and Bioinformatics
Jaypee University of Information Technology, Waknaghat, Solan–173234
Himachal Pradesh, INDIA

August, 2018

# Contents

**Chapter # 1:** INTRODUCTION

1. Background
    1.1 ILDs classification
    1.2 Clinical presentation of ILDs
    1.3 Clinical Management and Diagnostics of ILD
        1.3.1 Thoracic imaging:
        1.3.2 Genetic evaluation of ILDs:
    1.4 Problem Identification
    1.5 Objective of thesis
        1.5.1 Integrated Resource of radiological and clinical information for Interstitial Lung Diseases ILDs: ILD-DB
        1.5.2 Integrated Resource of genetic data for Interstitial Lung Diseases ILDs: ILDgenDB
        1.5.3 Identification of novel biomarkers, role of noncoding RNAs and differential pathways in ILD pathogenesis.
    1.6 Organisation of the thesis

**Chapter # 2:** INTEGRATED MEDICAL DATA RESOURCE AND META-ANALYSIS FOR DIFFERENTIAL PATTERNS IN INTERSTITIAL LUNG DISEASES (ILDS) FROM NORTH-WESTERN HIMALAYAN REGION: INTERSTITIAL LUNG DISEASE-DATABASE (ILD-DB**)**

    2.1 Background
    2.2 Material & Method
        2.2.1 Data collection and validation:
    2. 3. Result and Discussion
        2.3.1 ILD-DB resource
        2.3.2 Meta-analysis of ILD cases
    2.4 Conclusion

**Chapter # 3:** ILDgenDB: INTEGRATED GENETIC KNOWLEDGE RESOURCE FOR INTERSTITIAL LUNG DISEASES (ILDS)

    3.1 Background
    3.2 Material and Method
        3.2.1 Data collection and integration:
            i. Disease candidate genes (DCGs) identification
            ii. Gene normalization, and genomic & proteomic annotation
            iii. Analysis and integration of GO, miRNA and SNP with DCGs
            iv. Biomarkers identification
        3.2.2 Web implementation
    3.3 Results and Discussion
        3.3.1 ILDgenDB web application
            i. Disease gene network viewer

3.3.2 DCGs categorization and functional annotation, and their association with regulatory elements
i. Functional annotation of DCGs with GO analysis
ii. Pathways mapping of DCGs
3.3.3 Association of DCGs with regulatory elements
i. miRNA and DCGs association
ii. SNP and DCGs association
iii. SNP, miRNA and DCGs association
3.3.4 Biomarkers and their characterization
3.4 Conclusions

**Chapter # 4:** INTEGRATED ANALYSIS OF NON-CODING RNAS FOR THE IDENTIFICATION OF PROMISING BIOMARKERS IN INTERSTITIAL LUNG DISEASES
4.1 Background
4.2 Materials and Method
4.2.1 Identification of ncRNAs associated with ILDs
4.2.2 ncRNAs and pathways association analysis
i. Analysis of miRNAs associations with pathways and gene ontology
ii. Expression analysis of miRNAs
iii. Association analysis among miRNAs, lncRNAs and pathways
iv. Validations
4.3 Results and Discussion
4.3.1 ncRNAs, genes and pathways association analysis
i. 3.1.1 miR-pathways association analysis
ii. 3.1.2 lncRNA-pathways analysis
iii. 3.1.3 miRNA-lncRNA (ceRNA) interaction analysis
4. 4. Conclusion

**Chapter # 5:** CONCLUSION AND FUTURE WORK
5.1 Conclusion
5.2 Future Work

**Conclusion and future work**
**Reference**

# SUMMERY

Interstitial lung diseases (ILDs) are diverse group of ~200 acute and chronic pulmonary disorders characterized by a variable amount of inflammation and/or fibrosis. These diseases are associated with substantial morbidity and mortality due to misdiagnosis or delayed diagnosis. Clinical symptoms and radiological tests such as high resolution computed tomography (HRCT) and chest X-ray (CXR) hold more promises in developing countries due to their availability. Studies have also indicated the significant roles of genetic alterations of disease candidate genes (DCGs) in association with miRNAs and/or SNPs in ILDs pathogenesis. Regulatory elements such as microRNA (miRNAs), long-noncoding RNAs (lncRNAs), and the cross-regulation between miRNAs and lncRNAs towards competing endogenous RNAs (ceRNAs) activity are significant for disease diagnosis and monitoring. In first objective of this research work, focus has been given to the radiological assessment of ILDs to facilitate effective diagnosis through comparative analysis of clinical and radiological patterns. A comprehensive web-based resource of ILDs (ILD-DB) has been established for the radiological assessment of ILDs. Radiological (HRCT and CXR) and clinical profiles, and follow-up data from eleven ILD-subtypes including, a rare disease, pulmonary alveolar microlithiasis (PAM) were anonymized and incorporated. Dominant radiological patterns and the most reported diseases mainly from North-Western Himalayan region were reported in this study. All the data is freely accessible as public repository, ILD-DB, in healthcare domain (http://14.139.240.55/ilddb/). This study would be beneficial for characterizing new-onset ILDs, benchmarking of referral dataset, image processing algorithm development and radiological training. In the second objective, the first genetic knowledge resource, ILDgenDB, has been developed to provide ILDs genetic data and their integrated analyses for better understanding of disease pathogenesis and identification of diagnostics-based biomarkers. The catalog of DCGs and their association with various regulatory elements have been generated using an integrated bioinformatics workflow of databases searches and literature-mining. DCGs-miRNAs-SNPs and pathways association analyses responsible for the pathogenesis of twelve different classes of ILDs were also incorporated. Outcomes of analyses suggested the significant involvement of immune system and defense mechanisms in ILDs pathogenesis. This resource may potentially facilitate genetic-based disease monitoring and diagnosis. ILDgenDB is publically available at http://14.139.240.55/ildgendb/index.php. In the third objective, investigation of the role of ncRNAs as therapeutic/biomarker targets were implicated. The miRNAs and lncRNAs which are having higher genetic interactions with target genes/pathways/ceRNAs pairs were detected as promising candidates biomarkers. This study indicated the significant impact of ncRNAs and their target interactions. to provide the new directions for ILDs diagnosis and monitoring through non-invasive interventions. All the studies have been performed to address the current limitations in the diagnostics and therapeutics of the ILDs. This thesis work may assist researchers, pulmonologists and radiologists working for the improvement in the diagnostics and therapeutics of ILDs.

# CHAPTER # 1

## Introduction

Interstitial lung diseases (ILDs), also known as diffuse parenchymal lung diseases, encompass a diverse group of acute and chronic lung disorders with known and unknown causes that mainly involve the interstitium of the lungs [1]. Occasionally, these diseases also involve the alveoli, peripheral airways, vessels and pleura. According to the study on global burden of disease (1990–2013), these diseases are one of the ten causes of increased global deaths [2].

### 1.1 ILDs classification

These diseases may be acute, sub-acute or chronic, and may be reversible or irreversible. Many classifications exist for ILDs and no standardization can be developed. Thorasic societies over the world are working to create a consensus and created recommendation on classification and managements. The diseases are characterized by variable degrees of inflammation and fibrosis, and can be broadly classified into following four categories (Figure 1) [3, 4]:

i. Idiopathic interstitial pneumonias (IIPs) (e.g. idiopathic pulmonary fibrosis (IPF), non-specific interstitial pneumonia (NSIP), etc.).
ii. ILDs with a known cause such as drugs, environmental exposure (e.g. connective tissue disorder related-ILDs (CTD-ILD), silicosis, etc.);
iii. Granulomatous ILDs (e.g. sarcoidosis);
iv. Rare ILDs with well-defined clinical-pathological features (e.g. pulmonary langerhans cell histiocytosis and lymphangioleiomyomatosis)

IPF is an example of a chronic fibrosing type of IIP, characterized by a pathological and/or radiological usual interstitial pneumonia (UIP) pattern. The IPF occurs mainly in elderly people and adults, particularly in male smokers. Another variety of chronic fibrosing IIP is NSIP. Apart from being idiopathic, the NSIP pattern can also be seen in several other conditions like collagen vascular diseases (CVD), hypersensitivity pneumonitis (HP), and drug toxicity. Thus, multi-disciplinary discussions (MDD) involving clinicians, radiologists and pathologists are considered as the gold-standard in the diagnosis of ILD [4].

### 1.2 Clinical presentation of ILDs

Interstitial lung diseases (ILDs) are having very nonspecific and poorly distributed symptoms. The most common and primary symptom is shortness of breath as nearly all people with ILDs suffers with dyspnea which can get worsen over time. In most forms of interstitial lung disease, the shortness of breath develops slowly (over months). In interstitial pneumonias or acute interstitial pneumonitis, symptoms come on more rapidly (in hours or days). As due to stiffness in lung capacity get reduced

over time patient's often encounter breathlessness. This symptom called as hypoxia and it becomes worse over time, while daily activities such as walking, washing and dressing more difficult. Pulmonary hypertension is a characteristic of the latter stages of the disease and an important indicator of prognosis [5]. As the disease progresses, the patient may require supplementary oxygen [6].

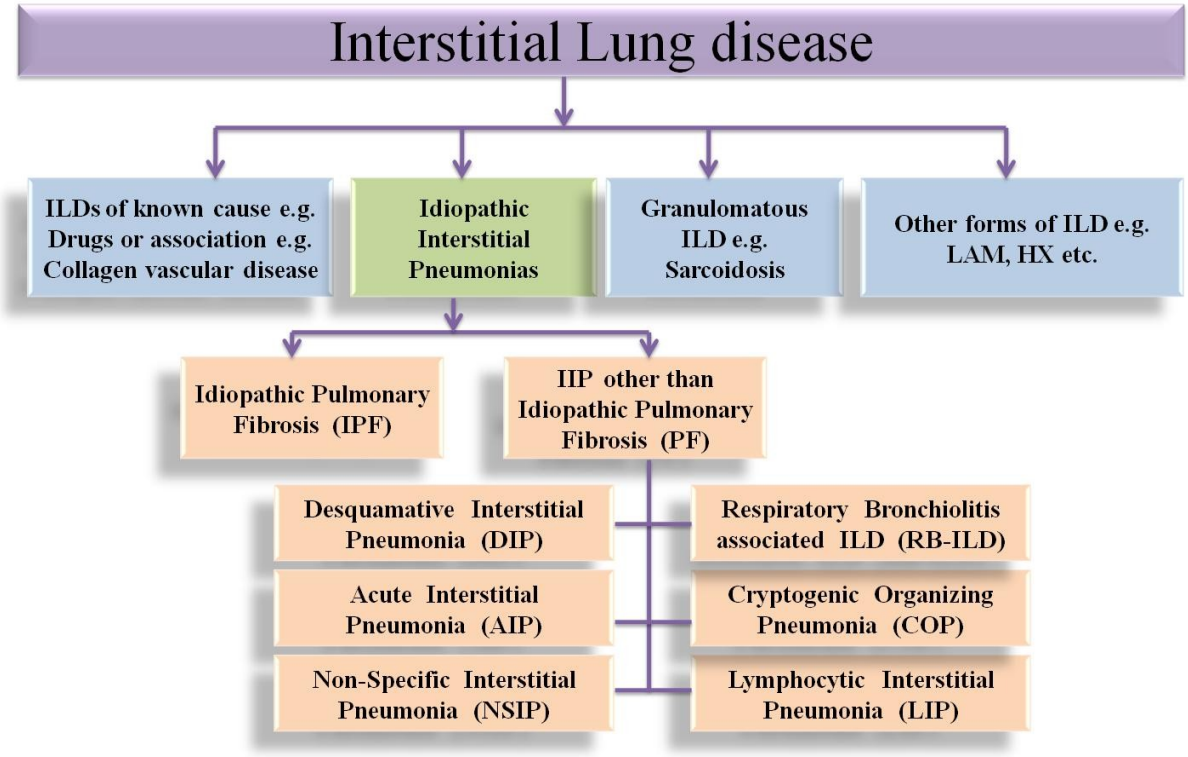

Interstitial Lung disease
ILDs of known cause e.g. Drugs or association e.g. Collagen vascular disease
Idiopathic Interstitial Pneumonias
Granulomatous ILD e.g. Sarcoidosis
Other forms of ILD e.g. LAM, HX etc.
Idiopathic Pulmonary Fibrosis (IPF)
IIP other than Idiopathic Pulmonary Fibrosis (PF)
Desquamative Interstitial Pneumonia (DIP)
Acute Interstitial Pneumonia (AIP)
Non-Specific Interstitial Pneumonia (NSIP)
Respiratory Bronchiolitis associated ILD (RB-ILD)
Cryptogenic Organizing Pneumonia (COP)
Lymphocytic Interstitial Pneumonia (LIP)

Figure 1: International Multidisciplinary Consensus Classification of ILDs by American Thoracic Society/European Respiratory Society

Few important symptoms of interstitial lung disease include:

i. Respiratory Symptoms: Shortness of breath, abnormal rapid breathing, reduced lung functions, and exercise intolerance are the most common complications of ILDs. With disease severity these problems worsen and lead to complete respiratory failure.
ii. Other Symptoms: Non-productive dry cough, unintentional weight loss, fatigue, joint pain and muscle ache is an important symptom of ILDs. In some form of ILDs sputum, mucus or blood in cough can be present.
iii. Velcro crackles: Adventitious breathing sounds have been identified as key features of early diagnosis of ILDs. These sounds can be easily diagnosed by physician during slow, deep breaths.
iv. Clubbing: Abnormal enlargement of the base of the fingernails and tips. Mostly occurs in disease advance stage.

v. Cyanosis: Another symptom of advance stage of ILDs. Due to decreased level of blood oxygen level patient's lips, skin and nails become bluish in colour.
vi. Genotype-Phenotype Correlation: Unknown, Mutation in few genes, abnormal or inadequate production of surface proteins.
vii. Radiological symptoms: Diffuse abnormalities on chest x-ray or CT scan.
viii. Inheritance Pattern: Autosomal Recessive; Autosomal Dominant

### 1.3 Clinical Management and Diagnostics of ILD

#### *1.3.1 Thoracic imaging*

ILDs-cases are very frequent in developing countries due to the use of tobacco, cooking by coal, and exposure to environmental hazards [7]. Incidence and mortality rates are found to be increased with age, exposure and other similar risk factors. Though abundance data on these diseases has not been available in India, five different studies from 1979 to 2016 indicated that the majority of ILDs patients suffered from idiopathic pulmonary fibrosis (IPF) and hypersensitivity pneumonitis (HP) [8]. The diagnosis of ILDs in developing countries is based on elicitation of thorough and comprehensive clinical history, including history of occupational and environmental exposures, drug intake, familial history, multi-systemic examination, chest X-ray (CXR) and high-resolution computed tomography (HRCT). In developing countries, there is always a looming possibility of under-diagnosis and under-reporting of ILDs [9, 10]. It may be due to the lack of acquaintance among doctors, or non-availability and/or non-affordability of various diagnostic modalities like high-resolution computed tomography (HRCT) scanning, bronchoscopy and video-assisted thoracoscopic surgery (VATS) [10, 11]. Another significant problem is the over-diagnosis of tuberculosis in ILD patients [12] [13]. Though surgical lung biopsy is the gold-standard in ILDs diagnosis, lack of specialist equipment and its affordability dissuade patients to explore this test [10]. The ATS/ERS/JRS/ALAT guidelines have emphasized on evidence-based ILD diagnosis and management and recommended that biopsy can be omitted in case of typical clinical and radiological patterns [3]. Therefore, the clinical and radiological tests (HRCT and chest X-ray (CXR)) hold more promises in developing countries. As overlapping and mimicking of the radiological patterns is frequent in ILD-subtypes (e.g. IPF, NSIP, etc.), diagnosis of any atypical or new onset presentation of patterns solely relies on assumptions or individual clinician's judgments which lead to empirical treatment and diagnostics [10]. Therefore, efficient ILD diagnosis needs to be based on integrated clinical history and physical examinations, radiological features and pathological reports of the patient [4, 14]. *Multidisciplinary discussion (MDD)* among pulmonologists, radiologists and pathologists is essential for the accurate diagnosis of these diseases [15], but it is generally not in practice in developing countries.

#### *1.3.2 Genetic evaluation of ILDs*

Studies have shown the evidence of genetic inheritances in familial silicosis, cystic fibrosis, bronchitis [15-17] and IPF [18]. Genetic studies performed on the animal model have given a clear insight into

the involvement of few genes on pathogenesis, prognosis and etiology of the ILDs [19]. Therefore, a comprehensive and comparative analysis of genes and other associated genetic data have potential to provide molecular diagnostic tools and candidate biomarkers. Recently, the focus of diagnostics has been shifted towards non-invasive techniques to reduce the adverse effects of radiations, frequent co-morbidities and biopsies. Ongoing studies have emphasized the crucial role of genetics in prognosis and etiology of ILDs [20] Molecular biomarkers for ILDs identification, response, and prognosis have the potential to improve current diagnostic approaches [21].

The involvement of noncoding-RNAs (ncRNAs) as regulatory elements, and their associated pathways and target genes were reported in pathogenesis and diagnosis of ILDs and other diseases [22-26]. These ncRNAs are mainly categorized as miRNAs, lncRNAs and ceRNAs. Studies have also indicated the regulatory role of ncRNAs in the immune system dysregulation, which is widely accepted as intervening factors in ILDs progression [27-30]. The ncRNAs are categorized as microRNAs (miRNAs) and long non-coding RNAs (lncRNAs) based on their transcript lengths [31]. These ncRNAs were reported as promising therapeutic/biomarkers targets to monitor the prognosis of idiopathic pulmonary fibrosis (IPF) and other ILDs [32, 33] . Effective analysis of disease-lncRNA association could benefit the ncRNA based therapeutic targets and diagnostic markers for different cancer and respiratory diseases [34, 35]. Additionally, the cross-regulation between lncRNAs and miRNAs towards competing endogenous RNA (ceRNA) activity were implicated in several biological processes, post-transcriptional regulation and diseases such as cancers sub-types and fibrotic lungs progression [26, 36, 37]. Accumulation of reported outcomes from ncRNAs-pathways and ceRNAs networks analysis led us to theorize that the ncRNAs influence molecular pathways, and can be key aspects in the etiology of ILDs. Therefore, these interactions could be functionally characterized and verified for their role as ILDs genetic biomarkers.

**1.4 Problem Identification**

As discussed earlier, the ILDs are very less known disease and most of the causes are still idiopathic. Also, ILD encompasses many different pathological processes so the treatment and regimen is different for each disease. An accurate and specific diagnosis is always required for successful management of ILDs. Many diseases have similar presentation and many have considerably varying symptoms. CXR and HRCT of the thorax can provide invaluable information that strongly supports a specific diagnosis. Many scientific committees including the European Commission's Radiation Protection Actions Committee (EUR16262), United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR), International Council of Radiation Protection (ICRP) and American College of Radiology (ACR) continuously raising serious concerns about the increasing radiation exposure from HRCT and CXRs. Studies are ongoing for radiation exposure and their provable biological effects and potential risks [38].

After 2008, ATS recommended use of surgical lung biopsy to confirm the histological diagnosis. Again these tests are critical with associated risk factors and very well trained practitioners are required to perform the task [39]. Sampling results of the body fluid indicated that the important role proteins secretion and their alteration in ILDs. These proteins are also present in bloodstream and serum even in small amount. These results suggest that genetic analysis of these protein, blood serum and associated genetic factors might represent a novel approach in the assessment of ILDs [40]. Although genetic testing is offered in daily practice in many countries, still no proper guidelines are available [40].

The successful establishment of these practices rely on availability of biological and clinical data from public resources. There are very less specialized databases/datasets available for ILDs in public domain. Thus, need of centralized databases for both genetic and clinical data can be clearly identified. In this thesis, we addressed these issues by developing two publically available data resources (ILD-DB and ILDgenDB) and performing meta-analysis of their data.

## 1.5 Objectives of the thesis

Following three objectives have been accomplished to address the current limitations.

### *1.5.1 Integrated Resource of radiological and clinical information for Interstitial Lung Diseases ILDs: ILD-DB*

The main objective was to create an integrated resource for ILDs referential data set that will be a collection of lung CT and X-ray images and other supporting information that can be used to develop decision support mechanism for ILD detection and determination of image lesions. This reference database could work as a uniform established system in order to provide a unified resource from point to point access, which could aids in diagnostics process in nationalised health services. The above objective has been fulfilled and detailed description of data collection, resource development and verification has been elaborated in Chapter 2.

### *1.5.2 Integrated Resource of genetic data for Interstitial Lung Diseases (ILDs): ILDgenDB*

ILDgenDB is a knowledgebase that integrates different types of functional and structural annotation and their interactions with sixteen different classes of interstitial Lung disease (ILDs). The interstitial lung disease candidate genes (DCGS) were curated from published articles. Manually mined genomic, proteomic and transcriptomic information, as well as orthologies, disease relationships, SNPs, gene function, and gene ontology analysis have been performed and added into one place. Keeping in view of the substantial utility of genetic data, this study discusses the major applications of ILDgenDB, a highly curated comprehensive knowledge resource, in ILDs diagnostics and therapeutics.

### *1.5.3 Identification of novel biomarkers, role of noncoding RNAs and differential pathways in ILD pathogenesis*

This study focused on comprehensive analyses of association among miRNAs-lncRNAs-pathway with ILDs genes, and their functional annotation using ontology analysis. The study also incorporated miRNAs-expression data analysis to dissect the role of individual miRNAs on ILD or its subtypes. This analysis provided systematic insights into the interactions among disease-associated ncRNAs, pathways and ceRNAs. This study would assist to identify the most promising ILDs-related ncRNAs and pathways as biomarker and therapeutics.

### 1.6 Organization of the thesis

The thesis has been organized into distinct four chapters as follows.

In this chapter (Chapter 1), the introductory part and literature's review of the work has been presented. In Chapter 2, a newly developed integrated resource of radiological and clinical information for interstitial lung diseases (ILD-DB) has been presented which can assist the diagnostics and radiological assistance of ILDs patients. In Chapter 3, first Integrated resource of genetic data for interstitial lung diseases (ILDgenDB) has been developed, and systematic analysis on literature curated disease candidate genes and their association with regulatory elements have been performed. In chapter 4, novel biomarkers, differential pathways and role of noncoding RNAs in ILD pathogenesis have been identified. Problems addressed by each objective are also mentioned in the conclusion section of each chapter.

# CHAPTER # 2

## Integrated Medical Data Resource and Meta-analysis for Differential Patterns in Interstitial Lung Diseases (ILDs) from North-Western Himalayan Region: Interstitial Lung Disease-Database (ILD-DB)

### 2.1 Background

Interstitial lung diseases (ILDs) are chronic disorders characterized by a variable amount of inflammation and/or fibrosis [39]. These diseases are associated with substantial morbidity and mortality due to misdiagnosis or delayed diagnosis. Clinical symptoms and radiological tests such as high resolution computed tomography (HRCT) and chest X-ray (CXR) hold more promises in developing countries due to their availability [41]. To facilitate effective diagnosis through comparative analysis of clinical and radiological patterns, a comprehensive web-based resource of ILDs (ILD-DB) has been established in North-Western Himalayan region. Radiological (HRCT and CXR) and clinical profiles, and follow-up data from eleven ILD-subtypes including, a rare disease, pulmonary alveolar microlithiasis (PAM) were anonymized and incorporated into this resource. Dominant radiological patterns and their distribution pointing to disease presentation were manually annotated and validated with the help of radiologists and included in the resource. Idiopathic pulmonary fibrosis (IPF) and connective tissue disease-associated ILDs (CTD-ILDs) were the most reported diseases in this study. Honey-combing and intra/inter lobular septal thickening patterns were predominant in the reported cases. All the data is freely accessible as public repository, ILD-DB, in healthcare domain (http://14.139.240.55/ilddb/). Availability of patients' clinical and symptoms data from the Himalayan regions could facilitate further insight into the etiology and management of ILDs. This study would be beneficial for characterize new-onset ILDs, benchmarking of referral dataset, image processing algorithm development and radiological training.

### 2.2 Material & Method

The data sets for ILD-DB are collected retrospectively from Indira Gandhi Medical College (IGMC), Shimla, India after the recommendations by expert pulmonologist and radiologist. All clinical data have been collected from standardised patient health record form. Radiological data such as Chest X-Ray (CXR), high resolution computed tomography (HRCT), and other pathological test data have been collected with robust data collection process. All these data and relevant information have been collected under certain protocols then best and referential data sets have been chosen and included in data repository for further training and testing phase.

### *2.2.1 Data collection and validation*

i. *Study Design:* All the data available in ILD-DB project was collected from ILD patients reported at IGMC Hospital, Shimla. Data have been taken and included in ILD-DB under the supervision of specialist pulmonologist and radiologist panel.

ii. *Data anonymization and Consent form*: A consent form is provided to patients in accordance with the regulatory and ethical requirements. All the patients who agreed to share their information have been taken. Patients had all the rights to discontinue/withdraw their participation at any point of time in the project.

iii. *Data annotation and validation:* DICOM (Digital Imaging and Communications in Medicine) image series were extracted from the picture archiving and communication system (PACS), and inbuilt graphical software was used to visualize and demarcate the region of interest (ROI) in entire HRCT slices. Dominant patterns (annotated ROIs) and their distribution pointing to disease presentation were annotated with the help of radiologists.

iv. *Integrated ILD resource development and data access:* Architecture and application of the proposed ILD-DB unified resource is provided in Figure 1. The HRCT-DICOM images, CXRs, ROIs and clinical data were stored in the database using MySQL server. HTML, PHP, and JavaScripts were used for graphical user interface (GUI) development.

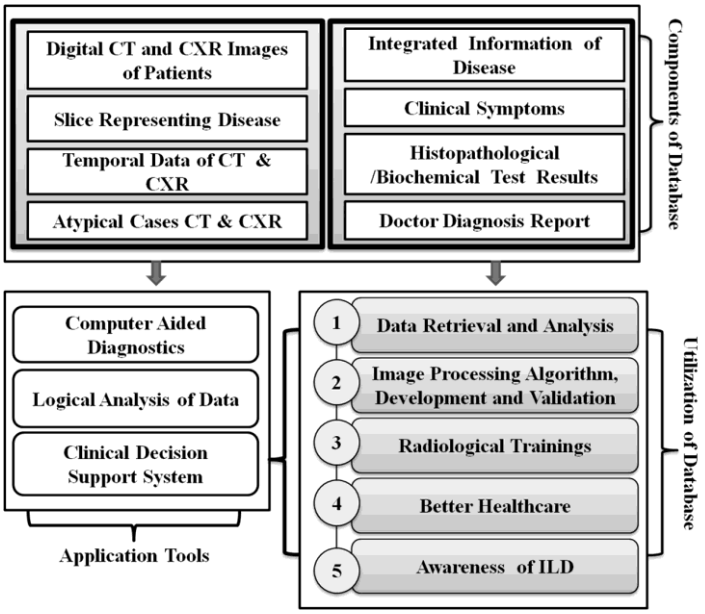



**Figure 1** Architecture and potential applications of the ILD-DB resource

## 2.3 Result and Discussion

This study presents a knowledge-resource and meta-analysis of different ILD-subtypes based on clinical and radiological patterns of patients. The meta-analysis presents a comparative analysis of symptoms and radiological patterns of this resource with earlier published studies from India and MedGift group [42]. All the data were systematically collected, analyzed and provided in the form of

EHR-enabled web resource named ILD-DB. Currently, this resource describes 238 cases with different lung diseases from different part of North-Western Himalayan region. Out of total 238 cases, following ILD sub-types consisting of 128 patients were reported: Combined pulmonary fibrosis and emphysema (CPEF); Pulmonary alveolar microlithiasis (PAM); Idiopathic interstitial pneumonia (IIP) (including Idiopathic pulmonary fibrosis (IPF) and Nonspecific interstitial pneumonia (NSIP), Sarcoidosis, Silicosis, Hypersensitivity pneumonitis (HP), Connective tissue disease related ILDs (CTD-ILDS) including- Progressive systemic sclerosis-associated ILDs (PSS-ILDs), Mixed connective tissue disease (MCTD-ILD), Systemic lupus erythematosus (SLE) and Rheumatoid arthritis-associated ILD (RA-ILD), and Unclassifiable-ILD (Table 1). Remaining cases (110) belong to obstructive lung diseases (i.e. chronic obstructive pulmonary disease (COPD) and bronchiectasis), infective diseases (i.e. TB and pneumonia) and other non-ILDs.

**Table 1** Clinical pattern analysis of different ILD-subtypes present in ILD-DB.

| S. No. | ILDs | Subgroups* | Number of Cases | Mean Age# | % Female# | % GERD* | % Clubbing * | Mean (FEV1/FVC) * | Mean FVC * | % of Smoking history |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | IIP | Unclassifiable IIP | 2 | 68 | 50 | 50 | 50 | NA | NA | 50 |
| 2 | | IPF | 40 | 61.7 (71.4) | 27.5 (50) | 35.5 | 46 | 93.7 | 64.32 | 38.4 |
| 3 | | NSIP | 15 | 52.2 (61.5) | 81.2 (50) | 40 | 50 | 89.5 | 65.2 | NO |
| 4 | ILDs of known cause | Silicosis | 10 | 51.7 | 0 | 40 | 40 | 77.3 | 73 | 40 |
| 5 | | PSS-ILD | 17 | 51.2 | 80 | 69.23 | 25 | 0.8 | 60.68 | NO |
| 6 | | RA-ILD | 5 | 54.6 | 60 | 75 | 0 | 84.9 | 55.26 | 33.3 |
| 7 | | MCTD | 3 | 43.6 | 100 | 100 | 66.6 | 80.1 | 67 | NO |
| 8 | | SLE | 3 | 48 | 33.3 | 50 | 0 | 62 | 62 | NO |
| 9 | Granulomatous ILDs | Sarcoidosis | 14 | 45.1 (48.5) | 78.5 (30) | 47 | 66.6 | 83.9 | 70.1 | 50 |
| 11 | | HP | 1 | 32 (65.3) | 0 (80.3) | 0 | 0 | NA | NA | NO |
| 12 | Other ILDs | PAM | 3 | 49 | 100 | NA | NA | NA | NA | NA |
| 10 | | CPFE | 11 | 69.9 | 10 | 44.4 | 50 | 90.3 | 90.33 | 74.9 |
| 13 | Unclassifiable-ILDs | | 5 | 64.8 | 80 | 60 | 33.3 | 85.5 | 63.3 | NO |

# Value in parentheses represents number in previous study [12]; * NA/NO: Clinical history was normal or test was not performed for the disease. GERD: Gastroesophageal reflux disease;

### *2.3.1 ILD-DB resource*

ILD-DB provides an extended interface for efficient and customized retrieval of data through "Query", "Browse" and "Download" options. The image (HRCT, CXR and ROIs) data is also available for download (Figure 1). Integrated information of the data resource can be accessed by

"disease name" using the drop-down menu and/or by providing radiological pattern information. Greater details on how to use, download and data accessibility options can be explored from ILD-DB tutorials accessible at http://14.139.240.55/ilddb/tutorial1.php.

### *2.3.2 Meta-analysis of ILD cases*

The distribution of clinical and radiological characteristics of observed patients are described in Table 1 and Table 2. Comparative analysis of radiological patterns and their distribution in different lung locations is presented in Table 2.

**Table 2** Comparative analysis of HRCT patterns in ILD-DB with other studies.

| SN | ILDs | Subgroups* | Depeursinge et al., (2012)# [42] | Singh et al., (2017) # [8] | ILD-DB# |
|---|---|---|---|---|---|
| 1 | IIP | Idiopathic pulmonary fibrosis (IPF) | Fibrosis, Bronchiectasis, GGO Peripheral, Subpleural, Basal, Posterior | UIP (Subpleural and basilar (P), Reticular abnormality, Honeycombing, | Honeycombing (F), Bilateral lower lobe (P), Intra and interlobular upper lobar septal thickening (F), Fibrosis, Cyst, minimal GGO (Bilateral Lower Lobe (P)), Traction bronchiectasis. |
| 2 | | Non-specific interstitial pneumonia (NSIP) | GGO, Consolidation, Reticulation, Fibrosis (uncommon) Peripheral, Basal | Basilar GGO (P) with or without sub-pleural sparing, Reticular abnormality (infrequent) | Intra and interlobular upper lobar septal thickening, GGO (Bilateral lower lobe, Right middle lobe), minimal Honeycombing (Right lower lobe), Nodules (Right middle lobe, Right lower lobe) |
| 3 | ILDs of known cause | Silicosis | NA | NA | Fibrosis or progressive massive fibrosis (F), Nodules (BUL), Alveolar opacities, Plural thickening, Intra and inter lobular upper lobar septal thickening, Bulla, Fibrocalcification |
| 4 | | Progressive systemic sclerosis associated ILDs (PSS-ILDs) | NA | Interstitial pneumonia patternscompatible with CTD: UIP, NSIP, LIP, OP | Intra and interlobular upper lobar septal thickening, GGO, Honeycombing and Emphysema |
| 5 | | ILD secondary to collagen vascular disease, MCTD | NA | | Intra and interlobular upper lobar septal thickening, GGO |
| 6 | | Systemic lupus erythematosus (SLE) | NA | | Intra and interlobular upper lobar septal thickening, Alveolar infiltrates (AI) |
| 7 | Granulomatous ILDs | Hypersensitivity pneumonitis (HP) | GGO, emphysema, Fibrosis, Diffuse | Upper lobe GGO (P), poorly defined centrilobular nodules; mosaic attenuation | Intra and interlobular upper lobar septal thickening & GGO (Diffuse) |
| 8 | | Sarcoidosis | Micronodules, Consolidation, Macronodules, Ground glass, Fibrosis (end-stage), Peribronchovascular, Subpleural, Peripheral | Upper lung perilymphatic distribution of nodules (P) with or without mediastinal or hilar lymphadenopathy | GGO (F), Mediastinal and hilar Lymph adenopathy (F), Tree in bud, Cyst, Intra and interlobular upper lobar septal thickening, Fibrosis, Traction bronchiectasis |
| 9 | Other ILDs | Combined Pulmonary Fibrosis and Emphysema (CPFE) | NA | NA | Fibrosis (F), Emphysema (F), Traction bronchiectasis, Honeycombing |

#GGO: Ground glass opacity F: Frequent pattern, P: Predominant location

*Desquamative interstitial pneumonia (DIP), acute interstitial pneumonia (AIP), cryptogenic organizing pneumonia (COP /BOOP), lymphocytic interstitial pneumonia (LIP) and langerhans cell histiocytosis (LCH) were excluded from this study as no cases were reported.

**2.4 Conclusion**

Integration of symptoms, pathological and imaging data to provide an efficient diagnosis and treatment is beyond reach of healthcare systems. Although high quality health services are available, majority of the populations cannot afford it. Therefore, efforts should be made to provide better diagnosis through state healthcare services. In this study, EHRs based web-repository was developed to share and access ILDs data. This resource gives due attention to security and human rights issues, and provides anonymized data. Comparative analysis of patterns from different geographical regions offers the opportunity for better understanding of ILDs risks and management. This study also presents a rare lung disease, PAM, with its radiological and clinical pattern analysis, which may update us about the current knowledge of this disease. This web accessible resource for ILDs data (radiological patterns reported in CXRs and HRCTs, relevant patient's history and lifestyle data) can facilitate computer-aided diagnostics through the development/evaluation of medical image processing algorithms. This study would be valuable to clinicians, especially for radiological training. This web resource, ILD-DB, is expected to work as a uniform resource for public awareness, and it can assist governments and international organizations for framing better policies.

# CHAPTER # 3

## ILDgenDB: Integrated Genetic Knowledge Resource for Interstitial Lung Diseases (ILDs)

### 3.1 Background

Interstitial lung diseases (ILDs) are a diverse group of ~200 acute and chronic pulmonary disorders which are characterized by variable amounts of inflammation, fibrosis, and architectural distortion with substantial morbidity and mortality [39]. Inaccurate and delayed diagnoses increase the risk, especially in developing countries [10]. Studies have indicated the significant roles of genetic alterations of disease candidate genes (DCGs) in association with miRNAs and/or SNPs in ILDs pathogenesis [43, 44] . Therefore, the first genetic knowledge resource, ILDgenDB, has been developed with an objective to provide ILDs genetic data and their integrated analyses for better understanding of disease pathogenesis and identification of diagnostics-based biomarkers. This resource contains literature-curated DCGs enriched with structural and functional annotations. The catalog of DCGs and their association with various regulatory elements have been generated using an integrated bioinformatics workflow of databases searches, literature-mining and DCGs-miRNAs-SNPs association analyses. Association analyses of miRNAs, SNPs and pathways responsible for the pathogenesis of twelve different classes of ILDs were also incorporated. Manually verified DCGs and their significant associations with SNPs, miRNAs and miR-polymorphisms were also provided. Experimentally confirmed miR-polymorphisms belong literature-mined and proposed biomarkers were identified, and incorporated in the resource. The ILDgenDB provided user-friendly browsing sections and extensive query-based information retrieval systems. Additionally, the resource also facilitates graphical presentation of the gene-SNP/miNAs interaction networks of the selected disease to aid to the data interpretation in an integrative way. Outcomes of analyses suggested the significant involvement of immune system and defense mechanisms in ILDs pathogenesis. The resource may potentially facilitate genetic-based disease monitoring and diagnosis. This resource is available at http://14.139.240.55/ildgendb/index.php.

### 3.2 Material and Method

#### *3.2.1 Data collection and integration*

*i. Disease candidate genes (DCGs) identification*

Schematic representations of different types of data incorporated in the knowledge resource are given in Figure 1. A systemic review of literature using advanced query was conducted in PubMed to identify the genes involved in ILDs. Genes associated with ILDs were manually curated and validated through published literature. Manually validated genes were enriched through information available in the gene-disease association databases such as KEGG, GAD, OMIM, GHR, CTD, DISEASE, and

GeneCards. The final list of validated genes associated with ILDs were considered as disease candidate genes (DCGs) and incorporated into ILDgenDB.

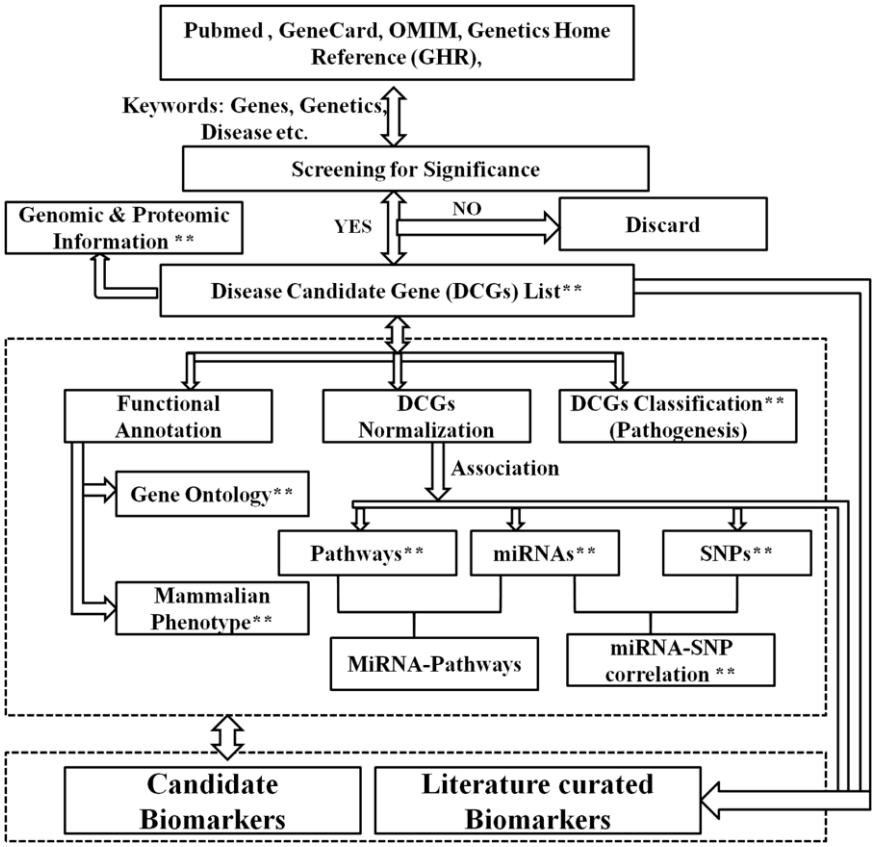



**Figure 1** Architecture of ILDgenDB knowledge resource. **The query terms that can be used to search the knowledge resource: name of each class of data such as gene, disease, category, biomarker, MiRNA AND/OR SNP, GO and phenotype may be used as query.

*ii.Gene normalization, and genomic & proteomic annotation*

For standardized recognition and cross-referencing of the DCGs, gene aliases, currently approved symbol, previous symbols and the standard IDs of different databases like Entrez, InterPro, Ensembl, HGNC, UniProt, OMIM, VEGA, MGI, etc., were included (Figure 1). For all DCGs, the important information such as functions, cryptogenic location, related nucleotide and amino acid sequences, pathways, mammalian phenotypes, mRNAs, domains, protein family, pathogenesis and other relevant information were also incorporated (Figure 2).

*iii. Analysis and integration of GO, miRNA and SNP with DCGs*

To characterize the functional importance of DCGs, the GO analysis was performed using AmiGO, Ensembl BioMart, DAVID and PANTHER. Only significant GO terms with a cut-off P-value of <=0.05 were considered, and the redundant data was removed. Although cut-off P-values 0.01 and 0.07 were also considered, the GO analysis with P-value<=0.05 provides optimal results. The DCGs

involvement with the biological processes, molecular functions and cellular component categories were added to the knowledge resources after verifying their association with diseases.

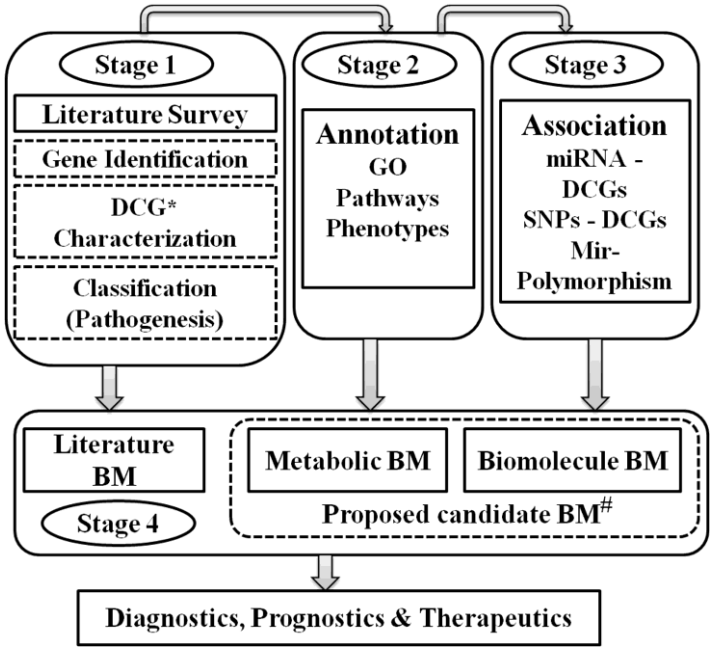



**Figure 2** Stage-wise workflow of ILDgenDB knowledge resource. *DCG: disease candidate gene; #BM: biomarkers: analysis of genetic data for the prediction of potential biomarkers.

Prediction of miRNAs interacting with mRNA (validated 299 DCGs) was performed using miRNA prediction tools and databases such as PITA, miRwalk, mirBase, starBase and targetScan. miRNAs were scanned and validated for their evidences of experimental verification and clinical significance with reference to the miRNA-DCGs association. After removal of ambiguous and duplicate values, genomic coordinates and mature sequences of the miRNAs were included in ILDgenDB.

Identification of SNP-DCGs association was performed using dbSNP. Only clinically significant SNPs and their annotations were used for filtrations of potential SNPs. The association among miRNA-SNPs-DCGs was also mined using the miRdSNP database. Only unique and clinically significant miRNA-SNPs-DCGs associations were incorporated into ILDgenDB. To validate association of these regulatory elements with ILDs, many miRNAs, SNPs and miRNA-SNPs-DCGs associations were also manually verified from published data.

*iv. Biomarkers identification*

The PubMed database was consecutively mined using a specific medical subject heading (MeSH) terms representing for twelve diseases and their established biomarkers. Different keywords (biomarkers, genetic testing, disease diagnostics, and molecular diagnostics, etc.) were used for retrieving data on reported biomarkers.

### *3.2.2 Web implementation*

The architecture of ILDgenDB knowledge resource is provided in Figure 1 which was implemented on MySQL server to provide fast executions of multiple queries defined by users. The first/user level was supported by JavaScript, PHP, HTML and cascading style sheets (CSS). This knowledge resource comprised of two different layers: data & application levels. The former stores annotation such as basic genomic and proteomic information and FASTA sequences of the DCGs. The application layer contains all the analysis outcomes performed on DCGs. These data are presented in an interactive manner to facilitate easy access and analysis.

## 3.3 Results and Discussion

Disease-specific genetic data resources containing genes, pathways, SNPs, mRNAs, miRNAs, etc., information assist in the identification of biomarkers, therapeutics targets and other molecular networks [45]. Several such data resources have been assisting in better disease monitoring and diagnosis [46]. However, a majority of resources containing these types of data are related to cancer or its subtype or diseases caused by bacteria and viruses [47]. ILDgenDB is the first knowledge resource containing reported DCGs, SNPs, miRNAs, biomarkers, different pathways and other relevant data related to twelve different types of ILDs (Figure 2, Stage 1). A literature survey was performed in PubMed and PMC to identify DCGs involved in ILDs. Genetic data and their role in disease pathogenesis were also deciphered. The sources or references about the DCG-disease association are also provided for every gene in the ILDgenDB resource.

### *3.3.1 ILDgenDB web application*

ILDgenDB provides an extended interface to explore its data through ten diverse query terms i.e. disease name, gene name, miRNA and/or SNP, etc. (Figure 1). For example, a user can select disease name “Pulmonary Fibrosis” from the drop-down list of “Browse Gene by Disease” option, then the name of all associated genes will be automatically displayed in the next drop down list, or any gene symbol can directly be entered to see the tabular results.

*i. Disease gene network viewer*

ILDgenDB provided an interactive visualization NetViz for different ILD subtypes. User can select disease of interest from drop down menu of different ILDs and select to view network either by “miRNA” or by “SNP”. Result page will redirect user to the interaction graph and assist to identify if any SNP/miRNA is associated with one or many potential DCGs. The interaction graph is a force directed plot and length of internodes shows the number of miRNA/SNPs interactions DCGs share with the ILDs. These associations could be further studied for their regulatory role in pathogenesis or as biomarkers.

### *3.3.2 DCGs categorization and functional annotation using GO and pathway mapping*

On the basis of involvement in ILDs pathogenesis, the DCGs were grouped into five categories such as therapeutic targets, biomarker and genetic testing, differential expression, mutations and others. On the basis of their associated pathways and interaction with other regulatory elements, 4% of DCGs such as IL-10, TGF-β, TNF, etc., are grouped as potential therapeutic targets. In total, 34% of the DCGs are categorized in biomarkers and genetic testing group, where 8% of them are established biomarkers and remaining 26% are under clinical trials or phenotypic testing. About 18% DCGs classified as others were found to have either indirect role or co-expressed with other key genes in ILDs pathogenesis. In total, 20% DCGs that shown up or down regulation in diseased versus control are grouped as differential expression. Approximately 24% genes have roles in mutations. These outcomes have shown that experimental validation of these DCGs can be useful for exposure induced ILDs.

*i. Functional annotation of DCGs with GO analysis*

To discover an association between DCGs with corresponding gene product's biology, functional and enrichment analyses of DCGs were performed. The biological processes, molecular functions and cellular components of all the DCGs were analyzed to identify the significant representation of GO terms. Outcomes of biological processes have shown that the maximum number of DCGs such as IL-4, IL-7, IL-8, IL-10, CCL- ligands, etc., are mapped with defense response (90), response to external stimulus (84) and immune system process (74). Identified cellular components for the majority of DCGs are extracellular region part (148) and extracellular space (112), and few are from the intrinsic component of the membrane (24). Results of molecular function have shown that the most of DCGs are mapped with cytokine activity (32), small molecule binding (12), chemokine receptor binding (11) and nucleoside binding (7). These results suggest the involvement of immune system and molecular binding in ILDs pathogenesis.

*ii. Pathways mapping of DCGs*

ILDs are progressive diseases that are associated with inflammation and lung fibrosis. ILDgenDB provides pathways mapping information of DCGs. These analyses indicate that the DCGs like TGF-β, SMAD, MMP-2 and MMP-9 are mapped with different pathways related to stimulation, signaling cascades and pro-fibrotic protein expression. Pathways analyses of the DCGs have suggested their significant involvement in the immune system (134 DCGs), innate immune system (91 DCGs), signal transduction (67 DCGs) and cytokine signaling (59 DCGs). Chemokine and cytokine signaling are other potential pathways where few important DCGs like IL9, IL10, TNF and CCL11 are involved. These outcomes suggest that the cytokine-cytokine receptor interactions may be targeted for disease diagnosis as these are actively participated in adaptive inflammatory host defense and the development and repair processes which are critical for the progression of ILDs. Targeting selective immune processes may give new insight to the novel therapeutics.

### *3.3.3 Association of DCGs with regulatory elements*

*i. miRNA and DCGs association*

To gain a global view of the molecular control networks, miRNA and DCGs association analysis was performed. The current analysis indicated that hsa-miR-335, hsa-miR-26b, hsa-let-7 and hsa-miR-30 target approximately 65, 34, 33 and 26 distinct DCGs, respectively. Differential expression of these target DCGs was reported and verified in many ILDs like IPF, systemic sclerosis, cystic fibrosis and acute lung injury [48, 49]. Furthermore, the DCGs that possibly interact with multiple miRNAs were also analyzed. DNAL1, MICA, FGF2, C3 and ITGA3 are the top five DCGs which are targeted by 175, 143, 99, 94 and 79 miRNAs, respectively. Other DCGs such as HMOX1, STAT3, etc., were proved to have significant roles in the pathogenesis of IPF [50], immune deficiency disorder and autoimmune ILDs [51]. The regulatory role of miRNAs (hsa-miR-335, hsa-miR-26b, hsa-let-7 and hsa-miR-30 etc.,) and altered biological pathways of top-ranked DCGs (DNAL1, MICA, FGF2, C3, ITGA3, etc.) could be explored as potential markers for ILDs diagnosis (Figure 2).

*ii. SNP and DCGs association*

SNPs associated with DCGs were identified using dbSNP and ensemble databases. Only validated SNP-DCGs associations were included into the ILDgenDB resource after taking into account 'Clinical Significance' terminology mentioned in dbSNP as filtration criterion. Genetic studies substantiated that SNP-gene associations can play a significant role in ILDs pathogenesis and progression [52]. In total, 2173 SNPs were reported in 15 genes (coding for 85 proteins) potentially associated with ILDs. Out of them, SNPs involved in DSP (860 SNPs) and SMAD3 (617 SNPs) were predicted which codes for 2 and 12 different proteins, respectively.

*iii. SNP, miRNA and DCGs association*

Ongoing cohort studies proved that miR-polymorphisms can regulate a variety of biological and pathological processes such as cell growth, differentiation, apoptosis and tumorigenesis [53]. A unique association analysis among SNP, miRNA and DCGs (miR-polymorphisms) was performed and incorporated in ILDgenDB. In total, 9170 miR-polymorphisms associations were identified for 91 DCGs. All these associations can directly be accessed using "miRNA-SNP correlation" option under "Disease" or "miRNAs AND/OR SNPs" as query option of the ILDgenDB resource. Out of 9170 miR-polymorphisms associations, 20 different genes: HMGA2, SPRED1, PLCG1, FKBP1A, FAM13A, DLG1, TSC1, NF1, FBN1, PIK3C2A, BDNF, FOXF1, MDGA2, CTGF, CXCL12, IL10, CAV1, NOG, FASLG and ITGA3 were found to have highest number of association. These 20 DCGs were associated with 5706 miR-polymorphisms indicated that these are from 229 different miRNAs, and the numbers of distinct SNPs in these miRNAs are 340. This outcome suggests that these miRNAs may further be experimentally validated to determine their potential role in ILD or its subtypes diagnosis and prognosis.

### *3.3.4 Biomarkers and their characterization*

In order to provide integrated information, ILDs biomarkers were collected from published literature and their annotations were manually verified and incorporated into "Biomarker" section of ILDgenDB resource. In total, 216 biomarkers were categorized as disease-specific protein (5), disease-specific (135), protein involved in pathogenic pathway (57), serum (40), disease-specific serum (7), protein (10), disease-specific miRNA (3) and gene (12). Majority of these biomarkers are used to determine whether the patient has ILD disease or not. However, very few biomarkers including KL6, SP-A, SP-D etc. are available to determine subtypes of ILDs. All these potential biomarkers require further validation (Figure. 2).

## 3.4 Conclusions

ILDgenDB is a centralized repertoire that hosts a wide range of ILDs data related to genes, proteins, biomarkers, etc., and their analyses. This resource contains 299 literature curated disease candidate genes (DCGs), and their association with SNPs, miRNA, pathways, biological processes, etc. The DCGs involvement in ILDs was verified by mining disease-related databases. The outcomes of the DCGs, GO, pathway and phenotypic analyses have indicated the potential role of the immune systems, molecular binding and inflammatory host defenses, and selective immune processes pathways in the pathogenesis of disease. Associations of DCGs with miRNA and SNP were performed which implicated many miRNAs and SNPs in expression alteration of these DCGs. Association analysis among SNP-miRNA-DCGs (miR-polymorphisms) was performed. A total of 229 miRNAs with 340 distinct SNPs were associated with 20 different ILDs genes. Some important DCGs such as cytokines, interleukins, surfactant proteins, etc., were also identified that may play a vital role in disease pathogenesis. Experimental validations of novel DCGs, miRNA, etc. can be valuable assets for the identification of new biomarkers. Provided datasets and their analyses would be helpful for researchers from diverse backgrounds working for the betterment of ILDs.

# CHAPTER # 4

## Integrated Analysis of Non-Coding RNAs for the Identification of Promising Biomarkers in Interstitial Lung Diseases

### 4.1 Background

Interstitial lung diseases (ILDs) are the group of chronic disorders with incomprehensible mechanisms. Dysregulation of immunological mechanisms may play significant role in ILDs progression [1]. Regulatory elements such as microRNA (miRNAs), long-noncoding RNAs (lncRNAs), and the cross-regulation between miRNAs and lncRNAs towards competing endogenous RNAs (ceRNAs) activity are significant for disease diagnosis and monitoring [31, 44, 54] . Systematic investigation and analyses of these regulatory elements can provide potential biomarkers for ILDs [55]. Current analysis explored the role of these regulatory elements as therapeutic/biomarker targets, and outlines the biological functions, differential expression and pathways association of the most significant ncRNAs [24]. In total, ten potential pathways related to ILDs and their associations with miRNAs and lncRNAs were identified. In the study, miR-pathways, lncRNAs-pathways and miRNA-lncRNA associations related to ILDs were implicated. The miRNAs and lncRNAs which are having higher genetic interactions with target genes/pathways/ceRNAs pairs were detected as promising candidates in our analysis. These results were also validated using literature-based knowledge. Analysis of miRNAs expression data indicated that few miRNAs may be used as ILDs or its subclass biomarkers. lncRNAs associated with miRNAs, and immunologically relevant pathways interactions were also proposed as biomarkers. The proposed miRNAs, lncRNAs and ceRNAs may be explored as potential biomarkers/therapeutic targets to assess ILDs pathogenesis and progression. The study indicated the significant impact of ncRNAs and their target interactions to provide the new directions for ILDs diagnosis and monitoring through non-invasive interventions.

### 4.2 Materials and Method

The procedure followed to identify potential ncRNA as therapeutic targets/biomarkers is given in Figure 1. Different steps performed for collection of data and their analysis are explained below.

#### *4.2.1 Identification of ncRNAs associated with ILDs*

Human miRNAs and lncRNAs were collected from the recent version of miRBase [56] and lncRNA databases (such as IRNdb [27] and lncRNAdb [57]), respectively. These ncRNAs were then mapped with their target ILD-genes available in ILDgenDB which provides published information on genes association with ILDs [58]. Gene-disease association and mode of pathogenesis were validated

with the help of standard web resources and databases [58]. Only ncRNAs mapped with target ILD-genes were selected for further analysis.

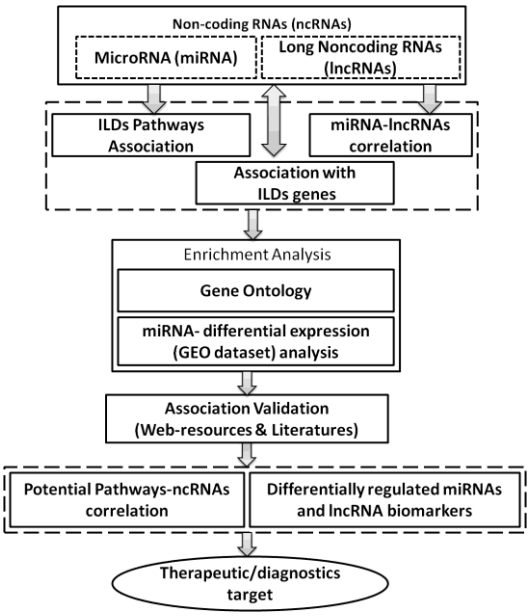



**Figure 1** Flow chart of methodology followed to identify ILDs' miRNA & lncRNA (ncRNA) as therapeutic target/biomarkers

### 4.2.2 ncRNAs and pathways association analysis

*i. Analysis of miRNAs associations with pathways and gene ontology*

The miRNAs association with pathways was performed using published literature and miRWalk2.0 comprehensive atlas [59]. Three pathways datasets from KEGG, Reactome, and WikiPathways were considered, and consistent annotations were used. A comprehensive list of experimentally verified miRNA associated with significant ILDs pathways was identified. In total, ten pathways were identified on the basis of their role in ILDs and interactions with miRNAs (Figure 2). Gene ontology (GO) analysis of these miRNAs were also performed (Figure.1). Only significant GO terms with a cut-off P-value of <0.05 were considered and redundant data was removed. Although cut-off P-values <0.01 and <0.07 were considered, the GO analysis with <0.05 provided the optimal results.

*ii. Expression analysis of miRNAs*

Gene expression omnibus (GEO) datasets were searched to select miRNAs's gene expression series. Keywords used for search were "miRNAs", "interstitial lung diseases", "idiopathic pulmonary fibrosis", etc. Total of six datasets (GSE75647, GSE21394, GSE81293, GSE8555, GSE27430, and GSE13316) were selected on the basis of availability and study design. GEO2R was used for the identification of differentially expressed genes by comparing groups of samples. Combination ranking of log fold change (LogFC) value (>=1.5 and <= -1.5) and P-Value (<=0.05) were applied to

determine the up- and down-regulated miRNAs, respectively. This ranking schema was defined by McCarthy et al. to define a biologically meaningful threshold [60]

*iii. Association analysis among miRNAs, lncRNAs and pathways*

The miRNAs-lncRNAs interactions were identified using starBase v2.0 [36]. LncRNAs-pathways associations were mapped with the help of target ILDs gene association, IRNdb and lncRNAdb. Those lncRNAs which are having significant association with ILD related miRNAs, pathways and genes were considered (Figure1).

*iv. Validations*

To determine the significance of these interactions in ILDs pathogenesis, the highly mapped terms related to the role of miRNAs-pathways, lncRNAs-pathways and miRNAs-lncRNAs associations were searched in Pubmed. A systematic review of literature was performed to evaluate the significant impact of ncRNAs–target gene networks in the genetics of ILDs. The advance query search function of PubMed was used for performing the search. Finally, ncRNAs were proposed as potential biomarkers on the basis of number of associations, differential expression and relevance determined through evidence from published literatures.

**4.3 Results and Discussion**

The ncRNAs are the potential modulator of genetic mechanism and pathogenesis for several diseases, including ILDs [60]. Systematic investigations for the role of ncRNAs in disease pathogenesis were performed mainly for cancers [31]. The associations among ncRNAs (miRNAs, lncRNAs and ceRNAs), ILDs pathways and genes were determined using in-house PERL scripts, bioinformatics tools, web resources and published literature from Pubmed. A total of 2701 miRNA-gene-pathways interactions, 669 lncRNAs-pathways and 4568 miRNA-lncRNA (ceRNAs) interactions were identified in the analyses. The miRNAs associated with multiple numbers of genes were recognized as potential biomarker/therapeutic targets in different studies [61, 62]. New therapeutic/diagnostics targets are also proposed on the basis of significance in pathways-ncRNAs correlation, differentially regulated miRNAs and lncRNAs-miRNAs associations (Figure1).

***4.3.1 ncRNAs, genes and pathways association analysis***

*i. miR-pathways association analysis*

Several potential ILDs-associated regulatory pathways and genes expression are altered by the miRNAs [9-10]. Manual curation and validation of such associations among miRNAs, ILDs pathways and target genes were performed (Figure 2). In total, 228 different pathways from KEGG, Reactome and WikiPathways were mapped to manually curated 127 genes from a total of 299 genes associated with ILDs reported in ILDgenDB resource.

Ten out of 228 different pathways were verified for their role in ILDs pathogenesis. The ncRNAs associated with these ten pathways were also identified (Table 2 and Figure 2).

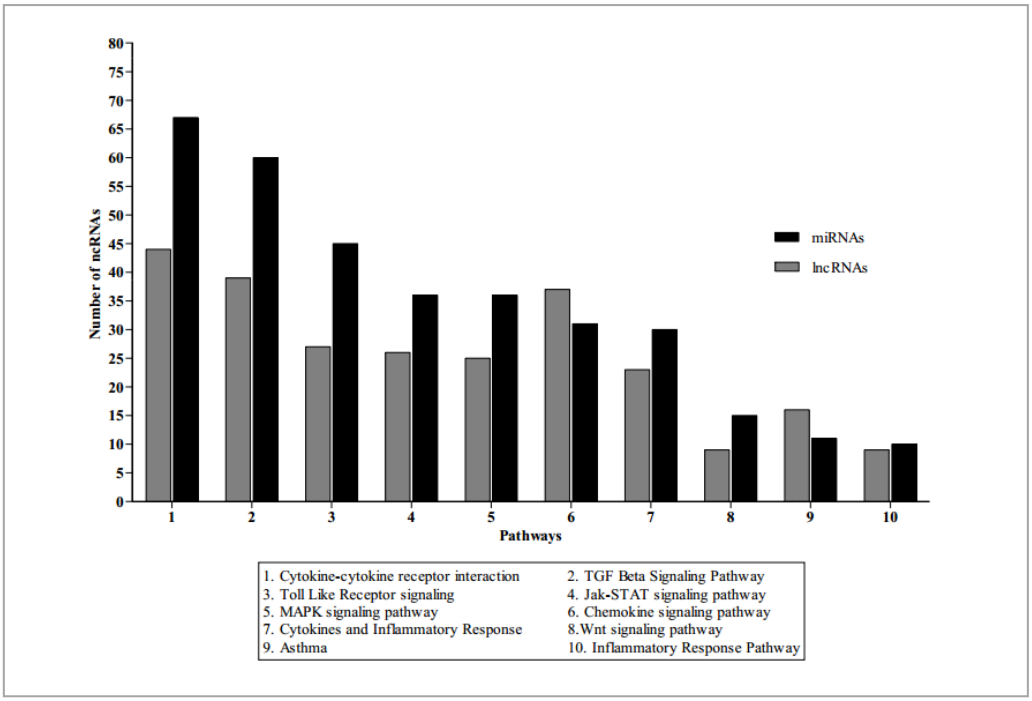



**Figure 2** Bar chart of miRNA and lncRNA associated with ten verified ILDs related pathways. Line and solid bars indicate the number of lncRNAs and miRNAs respectively.

*ii. lncRNA-pathways analysis*

Association analysis of lncRNA-Pathway identifies key lncRNAs related to different ILDs pathways (Figure 2) In total, 669 lncRNAs-pathways-genes interactions from 73 distinct lncRNAs, 196 pathways, and 91 ILDs genes were identified. Some potential lncRNAs mapped with maximum number of immunologically relevant pathways and ILDs genes are XIST (5), CTA-204B4.6 (5), MALAT1 (4), KCNQ1OT1, etc. The number of lncRNAs association with miRNA is presented in Table 2.

*iii. miRNA-lncRNA (ceRNA) interaction analysis*

The miRNA-lncRNA interaction (ceRNAs) serves as "sponge" to the former and prevents miRNAs-target binding [46]. Potential regulatory effect of the ceRNAs activity was proposed in IPF [26, 63]. In total, 4568 ceRNA interactions obtained from 108 miRNAs (involved in ILDs) and 909 lncRNAs were included in this study. The top ten ceRNA interactions associated with ten ILDs pathways and genes were proposed as biomarkers (Table 1 and Table 2). In Table 1, "Targeting ILD genes" are literature curated genes involved in ILDs progression or pathogenesis. "Number of associated lncRNAs" are specific lncRNAs that could be important for ceRNAs activity. "Number of associated pathways" is significant pathways manually verified with published literature for their role in ILDs.

**Table 1** Proposed potential miRNAs as therapeutic targets/biomarkers

| S. No. | miRNA | Targeting ILD genes | No. of associated lncRNAs | Number of associated Pathways[a] |
|---|---|---|---|---|
| **1** | hsa-miR-1 | AP3B1, BDNF, CCL2, EDN1, F2, HMOX1, IL6, IL8, ITGA3, TLR4 | 33 | 11 |
| **2** | hsa-miR-124-3p | BMP6, CAMP, CAV1, CCL2, DNAH5, EDN1, HMOX1, IL6, IL8, ITGA3, NME4, PGM1, PIK3C2A, PLA2G7, SERPINE1, SERPINH1 | 30 | 10 |
| **3** | hsa-miR-125b-5p | BMPR1B, IL1RN, ADM, BMPR1B, SLC9A3R2, STAT3, VDAC1, VDR | 26 | 8 |
| **4** | hsa-miR-155-5p | CAT, CCL2, EDN1, F5, FGF2, IFNGR1, IL6, IL8, NEU1, SLC9A3R2, SMAD3, STAT3, TTF1 | 46 | 10 |
| **5** | hsa-miR-21-5p | BMPR2, DERL1, FAS, FASLG, PIK3C2A, PLAT, SOD3, STAT3, TGF-β1, TLR4 | 19 | 10 |
| **6** | hsa-miR-26b-5p | ADM, BMP2, BMPR2, C3, CAV1, CCL2, CCL7, CXCL9, CXCR1, DERL1, FASLG, FGF23, GUSB, HMOX1, IFNG, IFNGR2, INHA, ITGA3, MMP8, PDGFB, SERPINH1, STX1A, TLR1 | 48 | 10 |
| **7** | hsa-miR-335-5p | ABCA3, ACE, ATF6, BMP2, CCR4, CCR7, CD14, CD27, CFTR, CLCA4, CXCL9, CXCR1, CXCR2, CYP2E1, FGF23, GC, GUCA2B, HMOX1, HSPG2, IL17A, IL1A, IL4, IL5, IL6, IL8, LTBP1, P2RY2, PDE5A, PIK3C2A, PLA2G7, PLAT, SERPINA1, SHH, SLC6A4, SMAD3, STX1A, TERT, THBD, TLR1, TLR2, TLR4, TNC, TREH, TSC1, VIP | 32 | 12 |
| **8** | hsa-miR-93-5p | BMPR2, DCTN4, STAT3, TOLLIP | 74 | 8 |
| **9** | hsa-miR-19a | BMPR2, TLR2, BMPR2, DERL1, TLR2 | - | 8 |
| **10** | hsa-miR-15a-5p | IFNG, PLA2G2D | 145 | 7 |

[a]Ten potential ILDs pathways: Asthma, Chemokine signaling pathway, Cytokine-cytokine receptor interaction, Cytokines and inflammatory response, Inflammatory response pathway, JAK/ STAT signaling pathway, MAPK signaling pathway, TGF-β signaling pathway, Toll-like receptor signaling, Wnt signaling pathway.

**Table 2** Proposed potential lncRNAs as therapeutic target/biomarker

| S. No. | lncRNAs | Association with No. of miRNAs | Association with No. of Pathways[a] | Association with No. of ILD-Genes |
|---|---|---|---|---|
| **1** | XIST | 87 | 5 | 45 |
| **2** | CTA-204B4.6 | 58 | 5 | 36 |
| **3** | MALAT1 | 42 | 4 | 27 |
| **4** | KCNQ1OT1 | 41 | 3 | 28 |
| **5** | ZNF518A | 40 | 4 | 25 |
| **6** | NEAT1 | 34 | 2 | 24 |
| **7** | SNHG16 | 33 | 4 | 24 |
| **8** | HCG18 | 32 | 4 | 22 |
| **9** | OIP5-AS1 | 29 | 2 | 17 |
| **10** | RP11-361F15.2 | 29 | 5 | 21 |

[a]All proposed lncRNAs biomarkers are associated to five pathways related to ILDs. These five pathways are Cytokine-cytokine receptor interaction, MAPK signaling, Toll-like receptor signaling, TGF-β signaling and Chemokine signaling pathway. Associations with number of miRNAs are the ILDs associated miRNAs and association with number of ILD-Genes validated with literature mining.

### 4.4 Conclusion

The study presents a comprehensive overview of the ncRNAs, ILDs-gens and pathways networks perturbed in ILDs and discusses key regulatory association. In total, twenty significant ncRNAs biomarkers/therapeutic targets (Table 1 & Table 2) and ten pathways (Figure 2) related to ILDs were identified. These findings suggest the role of inflammatory responses, different signaling pathways, and cytokine-cytokine receptor interactions in ILDs pathogenesis and diagnostics. Association of these pathways and biological processes are already established for ILD-genes. Furthermore, this study established that miRNAs and lncRNAs are also associated with these pathways. The findings enhance the understanding of these signatures in ILDs pathogenesis and progression. The results suggest that these ncRNAs and pathways could be new potential targets for ILDs therapeutic intervention. The study also indicated the role of lncRNAs and lncRNA-miRNAs (ceRNAs) interaction in disease process. The ncRNAs are proven to be better biomarker for disease monitoring and diagnosis. Many studies have been advocated that ncRNAs based disease monitoring could reduce approximately 30.2–46.5 % of unnecessary biopsies in patients. Studies also reported that ncRNAs biomarkers have higher specificity and sensitivity then conventional biomarkers, and ncRNAs detected in body fluids can work as the classical noninvasive biomarkers for cancers and other diseases [64]. The outcomes of this study expected to provide better diagnostic tool for ILDs monitoring and detection, and allow building ncRNAs, ILDs-genes and pathways based molecular model to identify, test and predict new therapeutic interventions.

# CHAPTER # 5

# CONCLUSION AND FUTURE WORK

## 6.1 Conclusion

Current work advances the knowledge of interstitial lung disease, radiological patterns, predominance of the disease and patterns, disease candidate gene identification, comprehensive enrichment analysis and candidate biomarker discovery. The outcomes of this study expected to provide better diagnostic tool for ILDs monitoring and detection, and allow building ncRNAs, ILDs-genes and pathways based molecular model to identify, test and predict new therapeutic interventions. Additionally, effort has been made to develop EHRs system to share and access these information considering security and human rights issues to serve for public health. This research is expected to offers veritable potentials, particularly for the developing countries, including web-based monitoring to provide personalized effective treatments and better understanding of disease condition.

## 6.2 Future Work

This research could lead us to advance many possibilities in advancement for radiology and genetics of ILDs. In the future, the two data resources namely ILD-DB and ILDgenDB can be used as training set and the data can be use for future application development and biomarker discovery. Following improvements can be incorporated in existing approach:

- Incorporation of machine learning for supervised classification and identification of ILD's specific patterns.
- Comparative analysis of radiological patterns to differentiate between ILDs and non-ILDs cases, and identify ILDs specific patterns among many ILDs classes.
- Disease candidate genes from ILDgenDB and ncRNAs biomarker can be used as training cases and similar gene expression patterns can be mined for ILDs specific genetic evolution for disease diagnosis and prognosis.

# Reference


1. Raghu G, Brown KK. Interstitial lung disease: clinical evaluation and keys to an accurate diagnosis. Clinics in Chest Medicine. 2004;25(3):409-19.
2. Vos T, Barber RM, Bell B, Bertozzi-Villa A, Biryukov S, Bolliger I, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. The Lancet. 2015;386(9995):743-800.
3. Raghu G, Rochwerg B, Zhang Y, Garcia CAC, Azuma A, Behr J, et al. An official ATS/ERS/JRS/ALAT clinical practice guideline: treatment of idiopathic pulmonary fibrosis. An update of the 2011 clinical practice guideline. American Journal of Respiratory and Critical Care Medicine. 2015;192(2):e3-e19.
4. Travis WD, Costabel U, Hansell DM, King Jr TE, Lynch DA, Nicholson AG, et al. An official American Thoracic Society/European Respiratory Society statement: update of the international multidisciplinary classification of the idiopathic interstitial pneumonias. American Journal of Respiratory and Critical Care Medicine. 2013;188(6):733-48.
5. Artham S, Somanath PR. Stage-Specific Effects of Hypoxia on Interstitial Lung Disease. Hypoxia and Human Diseases: InTech; 2017.
6. Lindell KO, Jacobs SS. Idiopathic Pulmonary Fibrosis: New guidelines for diagnosing and managing the disease demand a fresh approach to nursing care. AJN The American Journal of Nursing. 2003;103(4):32-42.
7. Johannson KA, Balmes JR, Collard HR. Air pollution exposure: a novel environmental risk factor for interstitial lung disease? Chest. 2015;147(4):1161-7.
8. Singh S, Collins BF, Sharma BB, Joshi JM, Talwar D, Katiyar S, et al. Interstitial lung disease in India. Results of a prospective registry. American Journal of Respiratory and Critical Care Medicine. 2017;195(6):801-13.
9. Singh V, Sharma BB. Laying the ground for research of interstitial lung disease in our country: ILD India registry. Lung India: official organ of Indian Chest Society. 2014;31(4):320.
10. Richeldi L, Rubin AS, Avdeev S, Udwadia ZF, Xu ZJ. Idiopathic pulmonary fibrosis in BRIC countries: the cases of Brazil, Russia, India, and China. BMC medicine. 2015;13(1):237.
11. Rosas IO, Dellaripa PF, Lederer DJ, Khanna D, Young LR, Martinez FJ. Interstitial lung disease: NHLBI workshop on the primary prevention of chronic lung diseases. Annals of the American Thoracic Society. 2014;11(Supplement 3):S169-S77.
12. Cosgrove GP, Bianchi P, Danese S, Lederer DJ. Barriers to timely diagnosis of interstitial lung disease in the real world: the INTENSITY survey. BMC pulmonary medicine. 2018;18(1):9.
13. Akhter N, Rizvi NA. Interstitial Lung Diseases misdiagnosed as tuberculosis. Pakistan journal of medical sciences. 2018;34(2):338.
14. Oikonomou A, Prassopoulos P. Mimics in chest disease: interstitial opacities. Insights into imaging. 2013;4(1):9-27.
15. Travis WD, King TE, Bateman ED, Lynch DA, Capron F, Center D, et al. American Thoracic Society/European Respiratory Society international multidisciplinary consensus classification of the idiopathic interstitial pneumonias. American Journal of Respiratory and Critical Care Medicine. 2002;165(2):277-304.
16. Donohue W, Laski B, Uchida I, Munn J. Familial fibrocystic pulmonary dysplasia and its relation to the Hamman-Rich syndrome. Pediatrics. 1959;24(5):786-813.
17. Feldmann A. [On familial disposition to silicosis]. Zentralblatt fur Arbeitsmedizin und Arbeitsschutz. 1960;10:229-33.
18. Bonanni PP, Frymoyer JW, Jacox RF. A family study of idiopathic pulmonary fibrosis: a possible dysproteinemic and genetically determined disease. The American journal of medicine. 1965;39(3):411-21.
19. Furukawa H, Oka S, Shimada K, Tsuchiya N, Tohma S. Genetics of interstitial lung disease: vol de nuit (night flight). Clinical Medicine Insights: Circulatory, Respiratory and Pulmonary Medicine. 2015;9:CCRPM. S23283.
20. Kitazawa H, Kure S. Interstitial Lung Disease in Childhood: Clinical and Genetic Aspects: Supplementary Issue: Current Developments in Interstitial Lung Disease. Clinical Medicine Insights: Circulatory, Respiratory and Pulmonary Medicine. 2015;9:CCRPM. S23282.
21. De Lauretis A, Renzoni EA. Molecular biomarkers in interstitial lung diseases. Mol Diagn Ther. 2014;18(5):505-22.
22. Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. Annual Review of Biochemistry. 2012;81:145-66.
23. Yarmishyn AA, Kurochkin IV. Long noncoding RNAs: a potential novel class of cancer biomarkers. Frontiers in genetics. 2015;6:145.

24. Chen T, He P, Tan Y, Xu D. Biomarker identification and pathway analysis of preeclampsia based on serum metabolomics. Biochemical and Biophysical Research Communications. 2017;485(1):119-25.
25. Cui H, Xie N, Thannickal VJ, Liu G. The code of non-coding RNAs in lung fibrosis. Cellular and Molecular Life Sciences. 2015;72(18):3507-19.
26. O'Reilly S. Epigenetics in fibrosis. Molecular Aspects of Medicine. 2017;54:89-102.
27. Ben-Hamo R, Efroni S. MicroRNA regulation of molecular pathways as a generic mechanism and as a core disease phenotype. Oncotarget. 2015;6(3):1594.
28. Denisenko E, Ho D, Tamgue O, Ozturk M, Suzuki H, Brombacher F, et al. IRNdb: the database of immunologically relevant non-coding RNAs. Database. 2016;2016.
29. Heward JA, Lindsay MA. Long non-coding RNAs in the regulation of the immune response. Trends in immunology. 2014;35(9):408-19.
30. Marsland B, Königshoff M, Saglani S, Eickelberg O. Immune system dysregulation in chronic lung disease. Eur Respiratory Soc; 2011.
31. Vencken SF, Greene CM, McKiernan PJ. Non-coding RNA as lung disease biomarkers. Thorax. 2015;70(5):501-3.
32. Chen X. Predicting lncRNA-disease associations and constructing lncRNA functional similarity network based on the information of miRNA. Scientific Reports. 2015;5:13186.
33. Kim Y-K. Extracellular microRNAs as biomarkers in human disease. Chonnam medical journal. 2015;51(2):51-7.
34. Martens-Uzunova ES, Böttcher R, Croce CM, Jenster G, Visakorpi T, Calin GA. Long noncoding RNA in prostate, bladder, and kidney cancer. European Urology. 2014;65(6):1140-51.
35. Schaukowitch K, Kim T-K. Emerging epigenetic mechanisms of long non-coding RNAs. Neuroscience. 2014;264:25-38.
36. Tay Y, Rinn J, Pandolfi PP. The multilayered complexity of ceRNA crosstalk and competition. Nature. 2014;505(7483):344.
37. Dai Q, Li J, Zhou K, Liang T. Competing endogenous RNA: A novel posttranscriptional regulatory dimension associated with the progression of cancer. Oncology letters. 2015;10(5):2683-90.
38. Kalra MK, Maher MM, Rizzo S, Kanarek D, Shephard J-AO. Radiation exposure from chest CT: issues and strategies. Journal of Korean medical science. 2004;19(2):159-66.
39. Raghu G, Collard HR, Egan JJ, Martinez FJ, Behr J, Brown KK, et al. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. American Journal of Respiratory and Critical Care Medicine. 2011;183(6):788-824.
40. Borie R, Kannengiesser C, Debray MP, Crestani B. The genetic diagnosis of interstitial lung disease: a need for an international consensus. American Journal of Respiratory and Critical Care Medicine. 2017;195(11):1538-9.
41. Raghu G, Mehta S. Interstitial lung disease (ILD) in India: Insights and lessons from the prospective, landmark ILD-India registry. Lung India: official organ of Indian Chest Society. 2016;33(6):589.
42. Depeursinge A, Vargas A, Platon A, Geissbuhler A, Poletti P-A, Müller H. Building a reference multimedia database for interstitial lung diseases. Computerized Medical Imaging and Graphics. 2012;36(3):227-38.
43. Brown D, Rahman M, Nana-Sinkam SP. MicroRNAs in respiratory disease. A clinician's overview. Annals of the American Thoracic Society. 2014;11(8):1277-85.
44. Cho J-H, Gelinas R, Wang K, Etheridge A, Piper MG, Batte K, et al. Systems biology of interstitial lung diseases: integration of mRNA and microRNA expression changes. BMC medical genomics. 2011;4(1):8.
45. McDermott JE, Wang J, Mitchell H, Webb-Robertson B-J, Hafen R, Ramey J, et al. Challenges in biomarker discovery: combining expert insights with statistical analysis of complex omics data. Expert opinion on medical diagnostics. 2013;7(1):37-51.
46. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. Journal of Clinical Epidemiology. 2005;58(4):323-37.
47. Pletscher-Frankild S, Pallejà A, Tsafou K, Binder JX, Jensen LJ. DISEASES: Text mining and data integration of disease–gene associations. Methods. 2015;74:83-9.
48. Sessa R, Hata A. Role of microRNAs in lung development and pulmonary diseases. Pulmonary circulation. 2013;3(2):315-28.
49. Christmann RB, Wooten A, Sampaio-Barros P, Borges CL, Carvalho CR, Kairalla RA, et al. miR-155 in the progression of lung fibrosis in systemic sclerosis. Arthritis research & therapy. 2016;18(1):155.
50. Selman M, Vicens V, Mendoza C, Montaño M, Ramos C, Romero Y, et al. Subsets of fibroblasts show resistance to apoptosis independently of their interstitial lung disease origin. Federation of American Societies for Experimental Biology; 2013.
51. Devine MS, Garcia CK. Genetic interstitial lung disease. Clinics in Chest Medicine. 2012;33(1):95-110.

52. Tochimoto A, Kawaguchi Y, Yamanaka H. Genetic susceptibility to interstitial lung disease associated with systemic sclerosis. Clinical Medicine Insights: Circulatory, Respiratory and Pulmonary Medicine. 2015;9:CCRPM. S23312.
53. George G, Mittal RD. MicroRNAs: Potential biomarkers in cancer. Indian Journal of Clinical Biochemistry. 2010;25(1):4-14.
54. Mathai SC, Danoff SK. Management of interstitial lung disease associated with connective tissue disease. BMJ: British Medical Journal (Online). 2016;352.
55. Kupczyk M, Kuna P. MicroRNAs—new biomarkers of respiratory tract diseases. Advances in Respiratory Medicine. 2014;82(2):183-90.
56. Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Research. 2013;42(D1):D68-D73.
57. Quek XC, Thomson DW, Maag JL, Bartonicek N, Signal B, Clark MB, et al. lncRNAdb v2. 0: expanding the reference database for functional long noncoding RNAs. Nucleic Acids Research. 2014;43(D1):D168-D73.
58. Mishra S, Shah MI, Sarkar M, Asati N, Rout C. ILDgenDB: integrated genetic knowledge resource for interstitial lung diseases (ILDs). Database. 2018;2018.
59. Dweep H, Gretz N. miRWalk2. 0: a comprehensive atlas of microRNA-target interactions. Nature methods. 2015;12(8):697.
60. McCarthy DJ, Smyth GK. Testing significance relative to a fold-change threshold is a TREAT. Bioinformatics. 2009;25(6):765-71.
61. Okada Y, Muramatsu T, Suita N, Kanai M, Kawakami E, Iotchkova V, et al. Significant impact of miRNA–target gene networks on genetics of human complex traits. Scientific Reports. 2016;6:22223.
62. Lu J, Clark AG. Impact of microRNA regulation on variation in human gene expression. Genome Research. 2012:gr. 132514.111.
63. Song X, Cao G, Jing L, Lin S, Wang X, Zhang J, et al. Analysing the relationship between lnc RNA and protein- coding gene and the role of lnc RNA as ce RNA in pulmonary fibrosis. Journal of cellular and molecular medicine. 2014;18(6):991-1003.
64. Li Y, Huang J, Guo M, Zuo X. MicroRNAs regulating signaling pathways: potential biomarkers in systemic sclerosis. Genomics, proteomics & bioinformatics. 2015;13(4):234-41.

## PUBLICATIONS

### Journals

1. ILDgenDB: Integrated genetic knowledge resource for interstitial lung disease (ILDs), S Mishra et al. (Oxrford, Databases). doi: 10.1093/database/bay053.
2. Integrated analysis of noncoding RNAs, differential expression, and pathway associations as potential biomarkers for interstitial lung diseases (ILDs), S Mishra et al. Gene Report, Elsevier. doi: 10.1016/j.genrep.2018.03.002
3. Integrated medical data resource and meta-analysis for differential patterns in interstitial lung diseases (ILDs) from north-western Himalayan region: interstitial lung disease-database (ILD-DB) S Mishra et al. (communicated)

### Conferences

1. Characterization of Radiological Imaging Biomarkers for the Identifications of Interstitial Lung Diseases (ILDs) for Developing Countries, S Mishra et al., 2016, International Conference on Parallel, Distributed and Grid Computing(PDGC)
2. Clinical and radiological decision support system prototype for characterization of interstitial lung disease (ILDs), S Mishra et al., International Conference on Signal Processing (ICSP-SATI 2016).
3. Computer Aided health care system for interstitial lung disease, S Mishra et al., R&D expo by IEEE-JUIT- 2016
4. Integration of medical Data for affordable healthcare, S Mishra et al. SciDataCon 2014.

### Databases

1. ILDgenDB- A comprehensive knowledgebase of the genes contributing to the etiology and pathogenesis of interstitial lung diseases (ILDs). Accessible at- http://14.139.240.55/ildgendb/index.php
2. ILD-DB- An integrated image and clinical profile resource of interstitial lung disease real time patients from IGMC, Shimla, India. Accessible at- http://14.139.240.55/ilddb/